

CAI
AK
-M19

IS Archives publiques
Canada

C

Canada

Government
Publications

Vol. 2 - No. 3
Fall 1984

Machine Readable Archives



BULLETIN

ISSN 0821-3658

This issue of the Bulletin is the first in a series devoted primarily to descriptions of data archives in Canada. The Social Science Data Archives at Carleton University is one of the oldest in Canada. In the first two articles, Wendy Watkins and Richard Gill have outlined the activities of archives and the project being undertaken to process the Canadian Gallup Poll data.

Carleton University Social Science Data Archives

Introduction

The Carleton University Social Science Data Archives (SSDA) acts as a repository for machine-readable social science data and provides a resource to facilitate data analysis.

It was first conceived in the mid-1960s, after the Department of Sociology and Anthropology began to acquire the Canadian Gallup Polls. The desirability of having a central repository for social science data had already been discussed, and the acquisition of the polls provided the necessary impetus for its establishment.

Since that time, SSDA's role has expanded. Rather than acting merely as a data repository, the Archives disseminates data; provides consultation to social science researchers regarding all aspects of quantitative analysis, from questionnaire design through to statistical analysis; and provides instruction in the use of the computer and statistical packages as research tools.

Holdings

SSDA receives data from several sources which may be loosely described in terms of five broad categories.

Direct Deposits

Principal investigators often deposit their data directly with the Archives. A recent example is the *Canadian Class Structure Project*, 1984, Myles, et al. This is the Canadian part of a multinational study of the labour process. It is based on a replication of Erik Olin Wright's study of the American class structure. Other countries participating in the study include Great Britain, Sweden, Finland, Norway, Australia and New Zealand. In addition, several nations are in the process of preparing proposals for similar studies.

The Canadian National Election Study, 1974, 1979, 1980, and *Quebec Referendum*, Clarke, et al., *The Canadian National Mobility Study*, 1974, Boyd, et al., and the *CBC/Southam/Carleton University School of Journalism Polls* covering the 1979, 1980 and 1984 federal elections as well as the recent Conservative and Liberal leadership conventions are

just a few examples of data sets deposited directly by principal investigators.

The Canadian Gallup Polls

SSDA is the Canadian repository for the Canadian Gallup Polls. It has the mandate to service all requests from Canadian sources for data and/or analysis.

The Carleton collection contains all existing data sets, beginning with the May 1945 poll. We receive new polls from the Canadian Institute of Public Opinion (CIPO) on a regular basis. The polls are currently being brought up to archival standards under a contract with the Machine Readable Archives Division of the Public Archives of Canada (MRAD/PAC). This project is described in another article in this issue.

Unfortunately, many of the earlier polls, conducted from 1945 to approximately 1965, have been lost. At that time, there was no perception that the polls might be of academic interest. Since SSDA became the official repository, however, the collection is virtually complete.

The Inter-University Consortium for Political and Social Research (ICPSR)

Through Carleton's membership, SSDA has access to all ICPSR resources. Over 200 data sets, ranging from American election studies to studies of international conflict, have been obtained from this source. Our membership also allows members of the Carleton community to participate in the Summer Program in Statistical Analysis and the Summer Workshop on Survey Analysis held annually at the University of Michigan in Ann Arbor.

Statistics Canada

Holdings from this source include User Summary tapes and Public Use Sample tapes from the 1971 and 1976 censuses, census tract data for Ottawa and Toronto from 1951 to the present, the 1981 Survey of Consumer Finances and several of the Labour Force surveys. We are currently in the process of ordering the CAN-SIM University Base and data from the 1981 census.

Other Sources

Other sources include data from various government departments, and other archives located in Canada, the United States, Europe and South America.

Services to the Carleton Community

Access to all SSDA resources is provided free of charge to all members of the Carleton community. While we primarily deal with Social Science users, the Archives has more recently been discovered by those in the Arts and Humanities.

We provide users with data for instructional and research purposes. Where we do not hold the appropriate data sets, we can order the data from the myriad of sources described above.

We also provide instructional services to students enrolled in courses dealing with quantitative analysis. In addition, we have sponsored several public presentations in our *Series on Secondary Analysis*. These focus on location of data, features of specific data sets, and techniques for analysis.

Services to External Users

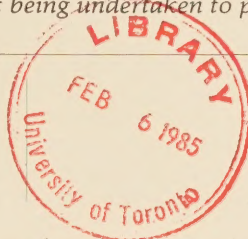
In addition to the provision of data sets and codebooks, services to those outside the Carleton community include locating data and conducting statistical analysis of SSDA or user-supplied data. This is done on a cost-recovery basis. For the most part, only data directly deposited with the Archives and Gallup Poll data may be made available to external users. The latter are available on a loan-contract basis only, and only with the written consent of the CIPO.

The charge structure for all services to external users (as of September 1984) is: staff time — \$35.00 per hour; computer time (internal rates) — cost recovery only; photocopying — \$0.25 per page; supplies — cost recovery; overhead — 30%. Requests for job estimates or for a listing of data available to external users are welcome.

Data Deposits

SSDA will accept machine-readable data from the behavioural sciences, as well as appropriate textual and historical machine-readable data for storage and dissemination on behalf of the depositors. Depositors may set reasonable restrictions on the use of their data (for example, restricting use to Carleton users only).

Cont'd on page 2





"Data Archives . . ."

cont'd from page 1

Often technical assistance and computer time to clean data or create codebooks can be provided. Depositing data with SSDA fulfils granting agencies' stipulations that sets be publicly available.

Archives Staff

SSDA currently has a staff of six. These include: Hyman Burshtyn, Director; Wendy Watkins, Data Archivist; Richard Gill, Project Manager, Canadian Gallup Poll Project; Laura Brewer, Research Assistant, Canadian Gallup Poll Project; Kathryn Chandler-Blockley, Documentalist, Canadian Gallup Poll Project; Dolores Leuty, Documentalist, Canadian Gallup Poll Project.

Richard Gill and Wendy Watkins

The Canadian Gallup Poll Project

A project is under way to clean and process nearly forty years of Canadian Gallup Poll survey data held by the Carleton University Social Science Data Archives (SSDA) as a result of a contract with the Machine Readable Archives Division of the Public Archives of Canada. SSDA is to produce standardized, verified raw data and SPSS system file versions of the polls, and a set of standardized, fully-documented codebooks.

The Gallup Polls are a valuable source of information about current and past public opinion, attitudes and values with respect to a wide range of social, cultural, economic and political issues in Canada. They are unique in that they are the longest-running continuous series of polls taken in Canada.

The appearance of standard demographic and political preference questions in the great majority of polls allows the analyst to trace trends or fluctuations over time, and to assess the impact of historical events and government policies on public opinion, and vice versa. For historians in all the social sciences there are valuable insights to be gleaned as well, from the popularity of horsemeat in 1951 (poll 212), to

opinions on capital punishment in 1960 (poll 280), to the role of the father in the family in 1976 (poll 390). Already the Gallup Polls have been used in studies on confidence in Canadian social institutions, attitudes and likely reactions to Quebec independence, attitudes toward multiculturalism, Canada-U.S. relations and numerous other topics.

Information on the following is included in almost all polls: province of residence, community size, marital status, mother tongue, level of education, occupation, income, religion, union membership, age, sex, present preferred political party and party voted for in the last federal election. The SSDA collection includes data sets for all existing polls, some 251 at present. The data for a number of early polls unfortunately was lost before SSDA began collecting them.

Much of the work for the Gallup Poll Project involves solving problems associated with data from the older polls. They often contain out-of-range values, mismatches between Gallup codebook information and the location of variables in the raw data themselves, labelling uncertainties and multiple-punched columns. The policy adopted for the project is to deal with these in such a way as to preserve the historical record while alerting the user to problems, rather than to make imputations and alter data in any way. Besides maximizing the detailed accuracy of the data sets, this approach also allows users to engage in methodological research, such as to trace the changing rate of appearance of wild codes for an historical study of the quality of survey data.

Along with the data, documentation and codebooks, a general background paper is being prepared describing, to the extent that reconstruction is possible, the changing practices of Gallup with respect to sampling, coding, and so on, and alerting users to ways in which the data should probably not be used.

Besides storing the Gallup material for posterity according to its standards and practices, the MRA will share with SSDA the right to disseminate the data for academic purposes. Canadian Gallup Poll Ltd. requires written approval for use of the data for non-academic purposes.

The project is expected to be completed by April 1985.

Richard Gill and Wendy Watkins

"A Canadian Adventure in Time"

In support of International Archive's Week, the Public Archives of Canada held an Open House on October 13 and 14. Over 2,200 visitors viewed the exhibits and participated in live demonstrations involving the conservation of pictures and the restoration of works of art. A sampling of many of the types of documents stored in the Archives were on display.

The Machine Readable Archives participated in the Open House displaying, through the use of a terminal, statistics from a one per cent sample of the 1871 Ontario census. These data result from a joint project between the Ontario Genealogical Society (OGS) and the Machine Readable Archives Division. Volunteers from the OGS are transcribing information from microfilm copies of the 1871 census to coding sheets. In turn, the MRA is computerizing the data in order to produce a machine-readable version of the nominal indexes. Indexes will be generated for each county throughout the province as well as an alphabetical listing of all individuals, by township, who lived in Ontario in 1871. Unfortunately, only a one per cent sample of the over 300,000 coded sheets was available for the Open House; the sample, however, provided interesting statistics on the country of origin, sex and occupation.

Also on display was the new slide and sound presentation which outlines the division's activities and provides pictures of the storage facilities. Many visitors were surprised to discover that the Archives was so involved with present technology.

Sue Gavrel

Notes

The Machine Readable Archives would like to disseminate information to the social science community on research projects involving the creation or use of machine-readable data. We would be delighted to accept brief articles on such projects which may be published in future issues of the *Bulletin*. For further information, please contact Sue Gavrel, Machine Readable Archives, 395 Wellington St., Ottawa, Ontario, K1A 0N3, (613) 993-7772.

des listes de codage. Le SSDA satisfait aux conditions des organismes de financement selon lesquelles les fichiers doivent être accessibles au public.

Personnel du SSDA

Le SSDA compte actuellement six employés : Hyman Burshlyn, directeur, Wendy Watkins, archiviste des données et, travaillant au projet sur les sondages Gallup, Richard Gill, gestionnaire, Laura Brewer, adjointe à la recherche, ainsi que Kathryn Chandler-Blockley et Dolores Leuty, documentalistes.

Richard Gill et Wendy Watkins

Projet sur les sondages Gallup au Canada

Le Social Science Data Archives (SSDA) de l'université Carleton a passé un contrat avec la Division des archives ordinolingués (AO) des Archives publiques du Canada afin qu'elle épure et traite les données des sondages Gallup couvrant près de quarante ans. Le SSDA produira des données brutes vérifiées et normalisées et des versions SPSS des fichiers, ainsi qu'une série de listes de codage normalisées et entièresment documentées.

Les sondages Gallup fournissent de précieuses informations sur l'opinion et l'attitude du public, aujourd'hui et dans le passé, à l'égard d'une foule de questions sociales, culturelles, économiques et politiques. Ils sont uniques du fait qu'ils constituent la plus longue série ininterrompue de sondages au Canada.

Comme la plupart des sondages comprennent des questions types d'ordre démographique et politique, l'analyste peut suivre les fluctuations ou l'évolution des tendances, et évaluer l'incidence des événements historiques et des politiques gouvernementales sur l'opinion publique et vice versa. Les historiens des sciences sociales peuvent eux aussi tirer des renseignements utiles de sondages tels que ceux concernant la popularité de la viande chevaline (n° 212, 1951), la peine de mort (n° 280, 1960) et le rôle du père de famille (n° 390, 1976).

Déjà on s'est servi de sondages Gallup pour mesurer la confiance des Canadiens dans leurs institutions sociales et connaître leur opinion sur l'indépendance du Québec, le multiculturalisme, les relations canado-américaines et de nombreux autres sujets.

Presque toutes les enquêtes comportent les questions suivantes : province de résidence, nombre de personnes habitant sous le même toit, état civil, langue maternelle, degré d'instruction, profession, revenu, religion, affiliation syndicale, âge, sexe, parti politique actuel du répondant et parti pour lequel il a voté à la dernière élection fédérale. Le SSDA possède les fichiers de presque tous les sondages existants, soit 251 jusqu'à maintenant. Malheureusement les données de certains anciens sondages ont été perdues avant que le SSDA ne commence sa collection.

Le projet Gallup consiste principalement à résoudre les problèmes liés aux anciens sondages de codage Gallup et à l'emplacement des variables dans lesquels on trouve souvent des valeurs hors fourchette, des incompatibilités entre les listes de codage Gallup et l'emplacement des variables dans les données brutes elles-mêmes, des étiquettes douteuses et des colonnes à perforations multiples. Il a été convenu de traiter ces problèmes de façon à préserver la signification des données tout en mettant l'utilisateur en garde, plutôt que de faire des imputations qui risquent de les fausser. En plus de maximiser la précision des fichiers, cette façon de procéder permet aux utilisateurs d'étudier les méthodes employées, par exemple de suivre les variations de fréquence des codes non pertinents pour évaluer la qualité des statistiques dans le temps.

Outre les données, la documentation et les listes de codage, on est en train de préparer un document général décrivant les différentes méthodes employées par Gallup pour l'échantillonnage, le codage, etc., dans la mesure où l'on pourra les reconstruire, et donnant des indications sur la façon d'utiliser les données.

Les AO stockeront les données Gallup suivant leurs normes et procédures pour les préserver. Elles auront le droit, tout comme le SSDA, de les diffuser à des fins de recherches universitaires. Ces données ne peuvent être utilisées à d'autres fins sans l'autorisation écrite de la firme Gallup.

Le projet devrait être terminé d'ici avril 1985.

Richard Gill et Wendy Watkins

Notes

Sue Gavrel

Archives étaient avancées sur le plan technique. Les Archives ordinolingués aimeraient diffuser de l'information aux personnes œuvrant en sciences sociales sur des projets touchant la création ou l'utilisation de données ordinolingués. Nous vous invitons à nous envoyer de brefs articles sur de tels projets pour les prochains numéros du *Bulletin*. Si vous désirez de plus amples informations, veuillez communiquer avec Sue Gavrel, Archives ordinolingués, 395 rue Wellington, Ottawa (Ontario), K1A 0N3, (613) 993-7772.

« Un voyage dans le temps »

Dans le cadre de la Semaine internationale des Archives, les Archives publiques du Canada ont tenu des journées d'accueil les 13 et 14 octobre. Plus de 2 200 personnes ont visité les expositions et assisté aux démonstrations liées à la restauration des œuvres d'art. Elles ont pu voir un échantillon des nombreux types de documents entreposés aux Archives.

Les Archives ordinolingués ont participé à ces journées d'accueil en présentant sur terrain un échantillon du recensement de 1871 en Ontario. Ces données résultent d'un projet mené conjointement par l'Ontario Genealogical Society (OGS) et la Division des archives ordinolingués. Des membres de l'OGS transcrivent les statistiques de 1871 enregistrées sur microfilm, puis les AO informatisent ces statistiques pour produire une version ordinolinguée des index de noms. Elles produiront ainsi des index pour tous les comités de la province ainsi qu'une liste alphabétique, par canton, de toutes les personnes qui vivaient en Ontario en 1871. Malheureusement seuls cent des quelques 300 000 feuilles codées ont pu être présentées lors des journées d'accueil; cet échantillon a tout de même fourni des statistiques intéressantes sur le pays d'origine, le sexe et la profession.

La division présentait aussi un nouveau diaporama sur ses activités et exposait des photographies de ses entrepôts. De nombreux visiteurs étaient étonnés de voir à quel point les Archives étaient avancées sur le plan technique.



Archives ordinologiques

Ce numéro est le premier d'une série consacrée principalement aux dépôts de données ordinologiques au Canada. Le Social Science Data Archives de l'université Carleton et Richard Gill nous parlent des activités de ce dépôt et du projet concernant les sondages Gallup.

Université Carleton Social Science Data Archives

Introduction

Le Social Science Data Archives (SSDA) de l'université Carleton sert de dépôt pour les données ordinologiques en sciences sociales et dispense de l'aide en matière d'analyse des

données. Il a été conçu au milieu des années soixante après que le Département de sociologie et d'anthropologie eut commencé à acquérir les sondages Gallup menés au Canada. On avait discuté déjà de l'opportunité de créer un dépôt central pour les données en sciences sociales, et l'acquisition des sondages a donné le coup d'envoi à ce projet.

Le SSDA a diversifié son rôle depuis. En plus de remplir les fonctions de dépôt, il diffuse des données, conseille les chercheurs en sciences sociales sur tous les aspects de l'analyse quantitative, depuis l'élaboration du questionnaire jusqu'à l'analyse statistique, et leur montre comment se servir des logiciels et des paquets-programmes.

Fonds

Le SSDA reçoit des données de diverses sources. Elles se divisent grosso modo en cinq catégories.

Dépôts directs

Les enquêteurs déposent souvent leurs données directement au SSDA. Un exemple récent est le *Canadian Class Structure Project*, 1984, par Myles et al., qui est la partie canadienne d'une étude multinationale sur la main-d'œuvre. Elle est basée sur une étude des classes aux États-Unis effectuée par Olin Wright. Les autres participants sont la Grande-Bretagne, la Suède, la Finlande, la Norvège, l'Australie et la Nouvelle-Zélande. En outre, plusieurs pays sont en train de préparer des études semblables.

Voici quelques exemples de fichiers déposés directement par des enquêteurs : Clarke et al., 1974, 1979, 1980, et le référendum québécois; Boyd et al., *The Canadian National Mobility Study*, 1974; et CBC/Southam/Carleton University School of Journalism Polls qui touche les élections fédérales de 1979, 1980 et 1984 ainsi

que les récents congrès d'investiture des Libéraux et des Conservateurs.

Sondages Gallup au Canada

Le SSDA est le dépositaire des sondages Gallup effectués au Canada. Il doit donc répondre à toutes les demandes d'origine canadienne concernant les données et les analyses de ces sondages.

La collection de Carleton renferme toutes les données recueillies depuis mai 1945 qui subsistent encore. Nous continuons d'en recevoir régulièrement de l'Institut canadien de l'opinion publique (ICOP). Les sondages sont actuellement rendus conformes aux normes archivistiques en vertu d'un contrat signé avec la Division des archives ordinologiques des Archives publiques du Canada (AO/APC). Ce projet est d'ailleurs décrit dans un autre article du présent Bulletin.

Malheureusement beaucoup des sondages effectués entre 1945 et 1965 ont été perdus. On ne croyait pas à l'époque qu'ils pouvaient intéresser les universitaires. Toutefois, depuis que le SSDA en est le dépositaire officiel, la collection est pour ainsi dire complète.

Carleton étant affilié à l'ICPSR, le SSDA and Social Research (ICPSR)

Statistique Canada

Ce ministère nous a envoyé des bandes sommaires et des bandes-échantillons à grande diffusion provenant des recensements de 1971 et de 1976, des données sur les secteurs de recensement à Ottawa et Toronto de 1951 à nos jours. L'enquête sur les finances des consommateurs (1981) et plusieurs enquêtes sur la

population active. Nous nous apprêtons à commander la base du CANSIM sur les universités ainsi que les données du recensement de 1981.

Services aux utilisateurs

de l'extérieur

Outre les fichiers et listes de codage que nous fournissons, les services dispensés aux utilisateurs de l'extérieur comprennent la localisation de données et l'analyse de données SSDA ou de données fournies par les utilisateurs eux-mêmes. Ces services sont offerts à prix coûtant. Généralement, les clients de l'extérieur ne peuvent obtenir que les données déposées directement au SSDA et les résultats des sondages Gallup. Ceux-ci ne peuvent être diffusés qu'en vertu d'un contrat et avec l'autorisation écrite de l'ICOP.

Voici quels étaient les tarifs en vigueur au mois de septembre 1984 : personnel, 35 \$ l'heure; temps machine (tarifs internes), prix coûtant; photocopies, 25 ¢ la page; fournitures, 30 %; frais généraux, 30 %. Nous répondons volontiers aux demandes d'estimation ou aux demandes de listes des données auxquelles les utilisateurs de l'extérieur peuvent avoir accès.

Dépôt de données

Le SSDA consentira à entreposer des données ordinologiques en sciences du comportement ainsi que des textes et des données historiques appropriées et de les diffuser au nom des déposants. Ceux-ci peuvent imposer des restrictions raisonnables sur l'utilisation de leurs données (p. ex. la limiter aux membres de Carleton). Souvent de l'aide technique et du temps machine sont fournis pour épurer les données ou créer